IN THE CLAIMS:

The text of all pending claims, (including withdrawn claims) is set forth below. Cancelled and not entered claims are indicated with claim number and status only. The claims as listed below show added text with underlining and deleted text with ~~strikethrough~~. The status of each claim is indicated with one of (original), (currently amended), (cancelled), (withdrawn), (new), (previously presented), or (not entered).

Please CANCEL claims 1-28, 31, 33, 44-48 and 52-57 and AMEND claims 30, 32, 50 and 51 in accordance with the following:

1-29. (cancelled)

30. (CURRENTLY AMENDED) A method for detecting similar documents using a computer, comprising ~~the steps of~~:

obtaining, using the computer, a document;

parsing, using the computer, the document to remove formatting and to obtain a token stream, the token stream comprising a plurality of tokens;

retaining, using the computer, only retained tokens in the token stream by using at least one token threshold;

reordering, using the computer, the retained tokens to obtain an arranged token stream;

processing, using the computer, in turn each retained token in the arranged token stream using a hash algorithm to obtain a single hash value for the document;

generating, using the computer, a document identifier for the document;

forming, using the computer, a single tuple for the document, the tuple comprising the document identifier for the document and the hash value for the document;

inserting, using the computer, the tuple for the document into a document storage tree, the document storage tree comprising a plurality of tuples, each tuple located at a bucket of the document storage tree, each tuple in the plurality of tuples representing one of a plurality of documents, each tuple in the plurality of tuples comprising a document identifier and a hash value; and

determining, using the computer, if the tuple for the document is co-located with another tuple at a same bucket in the document storage tree, thereby detecting if the document is similar to another document represented by the another tuple in the document storage tree.

31. (cancelled)

32. (currently amended) A computer-readable storage medium having software stored therein for causing a computer to perform ~~for performing~~ an operation using the method [[of]]in accordance with claim 30.

33-48. (cancelled)

49. (previously presented) A method as claimed in claim 30, wherein reordering is based on Unicode ordering.

50. (currently amended) A method for detecting similar documents using a computer, comprising ~~the steps of~~:

obtaining, using the computer, a document;

filtering, using the computer, the document to eliminate tokens based on parts of speech and obtain a filtered document;

generating, using the computer, a single tuple for the filtered document;

comparing, using the computer, the tuple for the filtered document with a document storage structure comprising a plurality of tuples, each tuple in the plurality of tuples representing one of a plurality of documents; and

determining, using the computer, if the tuple for the filtered document is clustered with another tuple in the document storage structure, thereby detecting if the document is similar to another document represented by the another tuple in the document storage structure.

51. (currently amended) ~~An apparatus~~A computer-readable storage medium having a program stored therein for causing a computer ~~for~~ to execute operations including detecting similar documents comprising:

~~means for~~ obtaining a document;

~~a filter to filter~~filtering the document to eliminate tokens based on parts of speech and obtain a filtered document;

~~a tuple unit to generate~~generating a single tuple for the filtered document;

~~a comparator to compare~~comparing the tuple for the filtered document with a document storage structure comprising a plurality of tuples, each tuple in the plurality of tuples representing one of a plurality of documents; and

~~a decision unit to determine~~determining if the tuple for the filtered document is clustered with another tuple in the document storage structure, based on the comparison, thereby detecting if the document is similar to another document represented by the another tuple in the document storage structure.

52-57. (cancelled)

58. (previously presented) A method as claimed in claim 30, wherein reordering is based on Unicode ordering.

59. (previously presented) A method as claimed in claim 30, wherein reordering is based on EBCDIC ordering.

60. (previously presented) A method as claimed in claim 30, wherein reordering is based on ASCII ordering.

61. (previously presented) A method as claimed in claim 30, wherein reordering is based on collection statistic measurements.

62. (previously presented) A method as claimed in claim 61, wherein collection statistic measurements are determined based on an inverse document frequency.